TRACEY PERSHOUSE
Associate Director
AECOM, Brisbane
tracey.pershouse@aecom.com

A MODELLER'S DILEMMA: OVERFITTING OR UNDERPERFORMING?

A transport model is devised with two primary requirements: replicating base year situations and forecasting future year implications. To fit base year situations to observed data, synthetic measures that cannot be explained by model inputs are often used, for example factoring trip generation results to match the observed number of trips. These overfitting measures reduce the model's responsiveness to the changes in model inputs, undermining the reliability of forecasting outcomes. In this paper we present the impact of applying overfitting measures on model outcomes through three case studies, each for one of the three model components: trip generation, trip distribution and mode choice. We show that whilst validation performance is improved, in terms of traffic volumes, trip length, and mode shares, this appearance is accompanied by adverse impacts on forecasting. To contrast, an alternative model is presented for each case, which removes the overfitting measures and includes additional explanatory variables. The alternative models have less desirable validation outcomes, but produce more responsive and sensible outcomes in forecasting. Our conclusion is that recognizing relative underperformance in the base year validation in exchange for more reasonable forecasting outcomes may be preferred to overfitting base year model to observed data.

1. Introduction

A transport model is devised with two primary requirements: replicating base year situations and forecasting future year implications. Testing the quality of a model is often focused on how well the model performs in validation testing and indeed, there are well established and accepted guidelines about testing validation performance. Tests to establish how reliably the model performs in forecasting, and although included in Guidelines like (UK Department of Transport, 2014) they are less widely used, reported or understood, and testing to establish the quality and reliability of the model's performance in forecasting receives less attention in practice. Robert Bain, after reviewing recent traffic forecasting reports, reported that 80-90% of the report was focused on the base year model whilst only 10-20% of the report focused on the future or forecasts (Bain, 2015).

In order to achieve an acceptable validation performance, measures are sometimes introduced which compromise the quality of the forecasting performance of the model. In this paper we illustrate how overfitting the base year can undermine the reliability of the forecasting outcomes. We present the impact of applying overfitting measures on model outcomes through three case studies, each for one of the three model components: trip generation, trip distribution and mode choice.

The case studies have been developed using common model forms in use in current Queensland strategic models. They are intended to illustrate problems and issues commonly encountered in modelling practice, although the case studies do not describe any actual model in use. The model forms used in the example application are well described in the literature and guidelines (for example Willumsen and Ortuzar, 2011)

1. Trip generation

1.1 Context

In strategic demand models, trip generation is almost always the first step, estimating the amount of trip making in the model area. There are two kinds of trip generation models used in traditional 4-step transport modelling processes: production models and attraction models. Trip production models estimate the number of home-based trips to and from zones where trip makers reside. Trip attraction models estimate the number of trips to and from each zone at the non-home end of the trip.

Linear regression is a common form, estimating the line of best fit between the number of trips made by a household or at zonal level, depending on a range of socio-demographic, land use and trip characteristics. Typical explanatory variables used to estimate the number of trips include variables related to person and household characteristics for trip production models and land use features for the attraction models. The literature recognizes that the estimated trip rates for trip attraction are generally less reliable than those for the production (Hall 2003). This is due to the reliability of land use data and also the complexity in the relationships between the trip attraction and land use characteristics. In practice major trip attractors such as major shopping centers are sometimes treated as 'special generators'. In this case study we demonstrate how the use of special generator to improve base year performance have compromised the model's forecasting capability and demonstrate an alternative model, which although it does not fit the base year data as well as the alternative, it does provide a better forecasting outcome.

1.2 Case study

Model Alternative A, flags a number of shopping centers as special generators. The number of Home Based Shopping (HBS) trips that are attracted to these locations related to the number of retail jobs, estimated using observed data on personal and household travel patterns sourced from household travel surveys. This estimate is then adjusted by an arbitrary, "special generator" adjustment factor chosen to match observed traffic count data. Whilst this adjustment factor improves the validation of the model, compared with observed traffic count data, it does not reflect underlying causes that necessitated this adjustment and therefore reduced the forecasting capability of the model to reflect future changes in land-use.

Model alternative B replaces the use of special generator adjustment factor with an additional explanatory variable, using land-use (in this case shopping area) density to influence trip attraction. The Alternative B model does not explain as much of the base year trip attractions as Alternative A, but it does improve the performance of the model in forecasting. The alternative B model is more plausibly sensitive to land-use changes than Alternative A.

1.2.1 Trip attraction general model form and parameters

The general form of the linear regression model for trip attraction is shown as follows in Equation (1)

$$A_i^p = \sum_{l \in L} r_l^p * l_i \qquad \text{Equation (1)}$$

where:

A_i^p is the number of trip attractions to zone i for trip purpose p
l_i is the measure of land use in zone i
r_l^p is the trip rate for the land use measure l and trip purpose p

Alternative A

The following processes are used to model the number of trip attractions for the HBS trips:

- Use the number of retail jobs as explanatory variables, and the trip rates per job were estimated from the Household Travel Survey (HTS) data as shown in the Table 1.
- Calculate the number of trips attractions to each zone based on the Equation (1), using the trip rate and the number of retail jobs in the zone. This produces the initial trip attraction results.
- Identify eight major shopping centers within the model area as 'special generators', and multiply the initial trip attractions to the 'special generators' by 2.5. This value of 2.5 is derived from the observed traffic counts at shopping centers with available count data, and adjusts for the difference between the initial trip attractions in step (2) and the observed count.

The use and adjustment of the 'special generator' is to reflect trip making behavior to large shopping centers offering a wide range of shopping, recreation and service facilities which tend to attract more trips per retail job compared with smaller, standalone, retail and service locations.

Table 1 Trip attraction coefficients for HBS trips, Alternative A

	Coefficients	t Stat
Intercept	0	#N/A
Retail (number of jobs)	8.54	27.85
Adjusted R square	0.912	

Alternative B

In this alternative model, retail job density[1] is introduced as an additional explanatory variable. This variables enables shopping centers to be differentiated from stand-alone, smaller retail outlets.

The trip attraction relationship is estimated similarly to Alterative A, and the resulting model is shown in Table 2. The density input is transformed to its logarithm value before being included in the linear regression.

Table 2 Trip attraction coefficients for HBS trips, Alternative B

	Coefficients	t Stat
Intercept	0	#N/A
Retail (number of jobs)	7.56	18.2
Ln(Retail job density, number of jobs per hectare)	484	3.19
Adjusted R square	0.921	

In both Alternative A and Alternative B models, the total number of HBS trip attractions sensibly increases as retail jobs increase. In Alternative B, the inclusion of density gives an additional incremental increase to the number of HBS trips, reflecting the increased attractiveness of larger shopping centers. In Alternative A, this effect is achieved by the use of a constant adjustment factor (in this case = 2.5).

Figure1 shows the form of Alternative B. It is analogous to trip rate relationships used in assessing in the assessment of traffic generating developments which typically show an increasing number of trips as retail floor area (and hence jobs) increase.

[1] It is calculated by dividing the number of retail jobs by the "commercial" area for the traffic zones with 50 or more retail jobs. It is assumed that the retail job density is zero for the zones with less than 50 retail jobs. This is because that some zones have a small number of retail jobs within a tiny area, whose high density level does not reflect their attractiveness to shopping trips.

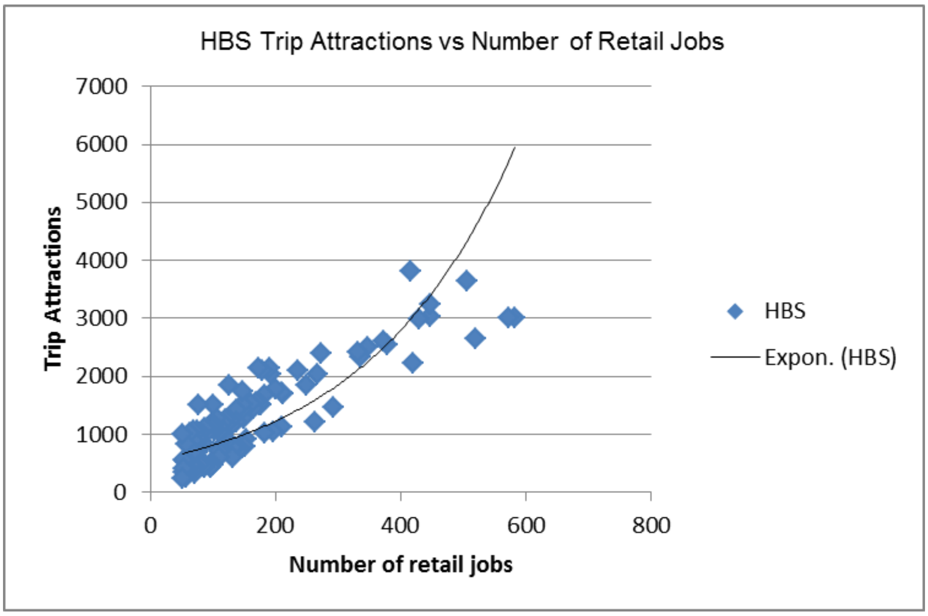

Figure 1 HBS trips vs zonal retail jobs

1.2.2 Validation Performance Compared

Table 3 presents the comparison between observed and modelled HBS trip attractions to major shopping centres. It should be noted that the observed number of trips are derived from the sampled HTS data instead of more reliable traffic counts, due to the lack of available counts surrounding the shopping centres.

It shows that the Alternative A produces good results for the largest shopping centre A, after the adjustment factor of 2.5 is applied based on the observed traffic counts. The Alternative B produces considerably worse validation results for the largest centres (A and B) but notably better outcomes for relatively smaller shopping centres (C to H).

Table 3 Comparison between observed and modelled HBS trips to major shopping centres

Shopping centre identifier	Observed ^ (Trips per day)	Alternative A	%diff (A - Obs)	Alternative B	%diff (B - Obs)
A	12,650	13,042	3%	9,496	-25%
B	11,533	7,059	-39%	3,905	-66%
C	5,215	1,677	-68%	1,898	-64%
D	3,364	8,156	142%	7,806	132%
E	3,157	7,549	139%	4,059	29%
F	2,936	1,442	-51%	1,369	-53%
G	2,676	5,219	95%	3,361	26%
H	1,460	6,604	352%	5,627	285%

^ HTS data

2.2.3 Forecasting Performance Compared

Figure 2 illustrates the increase of HBS trip attraction to the shopping centers from 2011 to 2031, in response to the growth of retail jobs. It shows that in the Alternative A, growth in jobs and growth in trip attractions are similar whilst in Alternative B, the number of trip attractions generally increases more rapidly than growth in jobs.

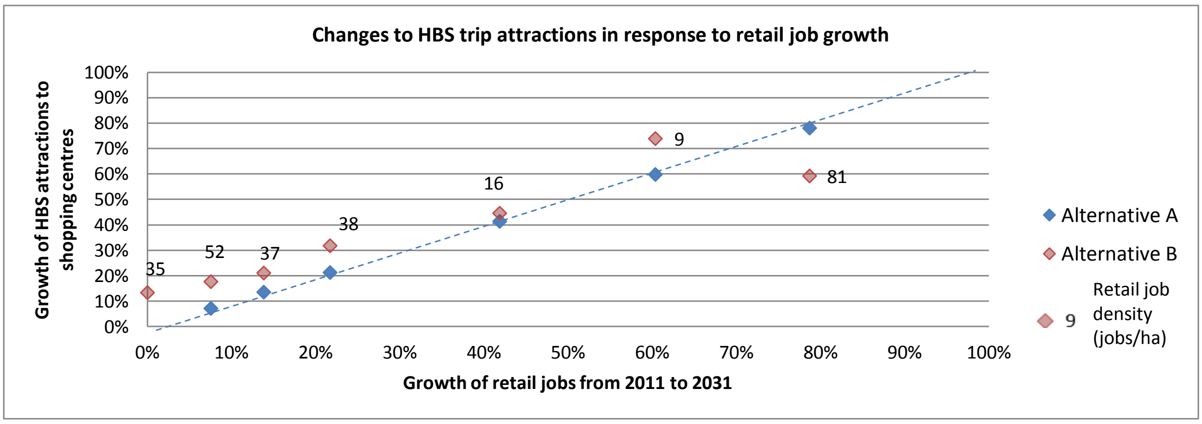

Figure 2 Changes of HBS trip attractions in response to the growth of retail jobs

1.2.3 Conclusions

Although Alternative B does not fit the observed data as well as Alternative A, Alternative B provides more plausible model outcomes. Trip attractions do not increase linearly with growth in jobs, rather they increase at a faster rate when the growth is added to a site with a relatively low job density, than when growth is added to a site with an already high job density. In other words, adding a large shopping centre to a low job density site will add relatively more trips, than increasing the jobs density at an already large shopping centre. Moreover by removing the need for a special generator adjustment factor, there is no need to specially identify zones as special generators.

2. Trip distribution

2.1 Context

A common form for trip distribution models is the gravity model which estimates the number of trips from a production zone attracted to each attraction zone, based on the relative cost of travelling between each zone and on the relative "attractiveness" of each attraction zone. The relative attractive - ness of each zone is usually related to land-use characteristics of the zone.

There are a number of practical issues encountered when calibrating gravity model, using household travel survey data. The sample sizes of HTS data typically result in very sparse matrices, where many zone to zone combinations have zero observed trips. This zero observation may be a result of small sample size, or there may be no trips made between these zones in reality (because there are no productions from A and no attractions to B so no trips between A and B), or there may be no trips because the cost of travelling between zones A and B is very high.

Excluding the zero trip ODs from the calibration produces a better fit between observed and modeled trip lengths. However, by excluding OD pairs with zero observed trips, the model is only applied to with non-zero observed trips and this brings the resulting modeled trip length distribution much closer to the observed data. However this close fit, is only illusory. In model application the gravity model is applied to all origin-destination zones, and therefore many small non-zero trips are estimated in model application.

2.2 Case study

2.2.1 Distribution model general form and parameters

The gravity model is described by the formulation below:

$$T_{ij}^{k} = \frac{\alpha_i P_i^k \beta_j A_j^k f^k(c_{ij})}{\sum_{j=1}^{n} A_j^k f^k(c_{ij})} \qquad \text{Equation (2)}$$

Where

T_{ij}^{k} = production to attraction trips from production zone i to attraction zone j for demand segment k

α_i, β_j = calibrated model coefficients

P_i^k = total production trips for segment k of production zone i

A_i^k = total attraction trips for segment k of production zone i

$f^k(c_{ij})$

= the impedance function of the travel cost (in minutes)between production zone i and attraction sone j for demand segment k

And the impedance function is as below:

$$f(C_{ij}) = C_{ij}^{x1} x \exp(-x_2 C_{ij}) \qquad \text{Equation (3)}$$

Where

$f(C_{ij}) = \textit{impedance factor between production zone i and attraction zone j}$

$\textit{Cij generalise cost between i and j}$

$x1 \textit{ and } x2 \textit{ are coefficients to be calibrated}$

Alternative A

In this Alternative, only OD pairs with observed trips are included in the dataset used for model estimation and only zones with observed productions or attractions. The resulting parameters for the impedance function as shown in the Table 4.

Table 4 Impedance function coefficients, Alternative A

	X1	X2
Coefficient	0.258	0.023

Alternative B

In this alternative, costs for all origin-destination pairs are included in the dataset used for model estimation, and the modeled production and attraction totals are used, replacing the observed trip end totals. All the sector to sector movements are included in the calibration, including those with zero trips. SA2 sectors are used here. This is to overcome the limitation of sparse HTS data that leads to a large number of origin-destination movements with zero trips. By aggregating the trips from zone to sector level, the sector to sector movements with zero observed trips are less likely due to the small sample size in HTS, and are more likely related to the amount of trip generations and travel costs. This allows including the zero trip movements as part of the observations used for the calibration. The travel costs between two sectors are derived from the average cost of all the movements between the two sectors weighted by the observed trips of each movement. The resulting parameters for the impedance function as shown in the Table 4.

Table 5 Impedance function coefficients, Alternative B

	X1	X2
Coefficient	0.719	0.100

2.2.2 Validation Performance Compared

Two types of trip length comparisons are used to assess the performance of the trip distribution models. In the first comparison, the observed trip productions and attractions (P & As) are used. The distribution of the resulting estimated trip lengths are compared with the observed trip lengths recorded in the HTS dataset. This comparison shows that given a set of productions and attractions how well does the model replicate observed trip lengths. Figure 4 presents the comparison of trip length distributions. It shows that the Alternative A produces a significantly better fit to the observed data that the Alternative B. As Alternative B uses modeled P&As to calibrate the model parameters, not observed, this difference in fit is expected.

However, in actual model applications, modeled productions and attraction totals are input to the distribution model, not observed trip ends. So in this second comparison, modeled P&As are used are used to compare the resulting distribution of trip lengths with the observed trip lengths. Figure 5 shows the trip length comparison. The Alternative A produces significantly longer average trip lengths compared with the HTS data, when the observed P&As are replaced with modelled values for producing outturn trip lengths. In contrast, the Alternative B produces good fitness to the observed trip lengths.

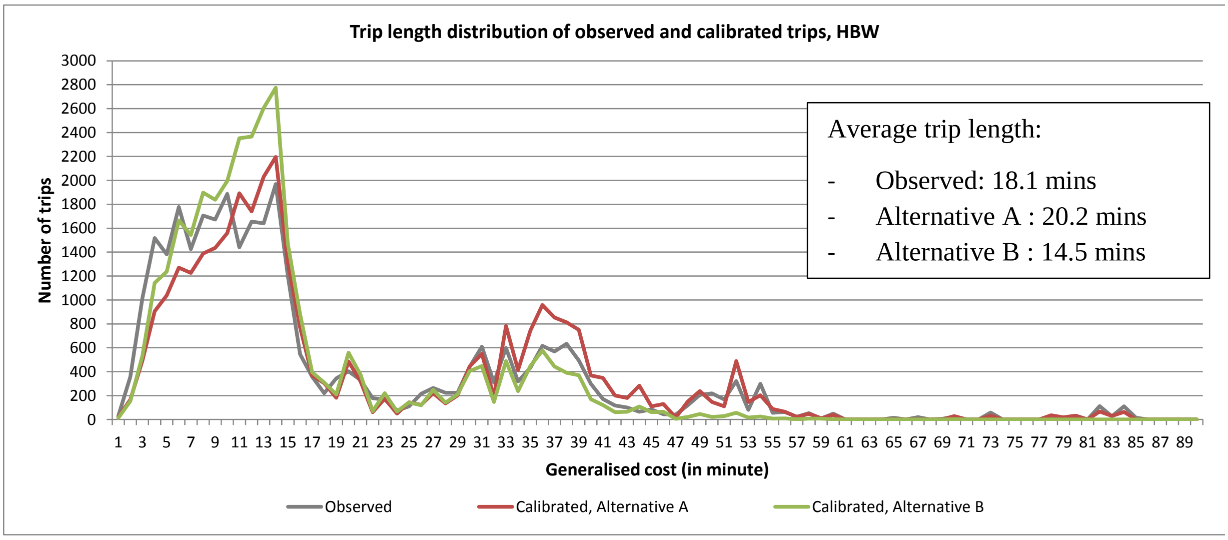

Figure 4 – Comparison of trip length distribution, Alternative A and Alternative B using observed P&A

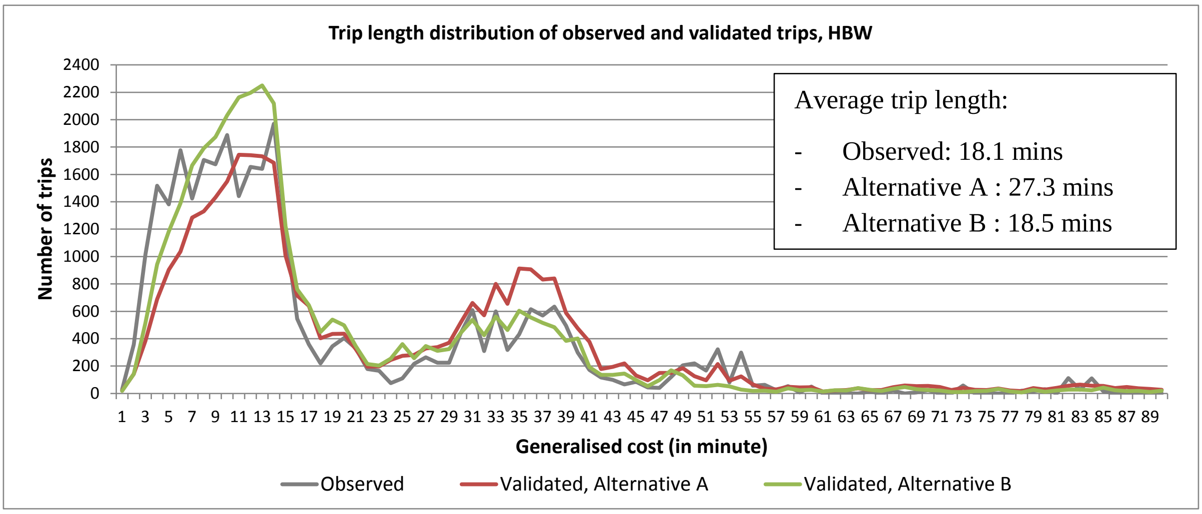

Figure 5 Comparison of trip length distribution for model validation. Alternative A and Alternative B using modeled P&As.

2.2.3 Forecasting Performance Compared

Table 4 compares the modelled elasticity of changes in trip length in response to cost increase between the two alternatives. The Alternative B produces a significantly greater reduction in trip lengths when travel cost increases compared with Alternative A. The elasticity declines (in absolute value) gradually in Alternative B from -0.4 to -0.2, as the costs increases from +20% to +100%. This diminishing impact of increasing cost on trip length realistically reflects factors other than the spatial separation between trip production and attractions. The more realistic sensitivities produced by Alternative B is likely to produce more plausible forecasting outcomes.

Table 4 Elasticity of changes in trip length in response to cost changes (Home based work trips)

Increase of generalised costs	Alternative A		Alternative B	
	change to average trip length	Elasticity	change to average trip length	Elasticity
+20%	-4%	-0.18	-8%	-0.40
+40%	-7%	-0.17	-13%	-0.33
+60%	-10%	-0.17	-16%	-0.27
+80%	-13%	-0.16	-18%	-0.23
+100%	-16%	-0.16	-20%	-0.20

3. Trip distribution and mode choice

3.1 Context

In many four-step models the mode choice occurs after the trip distribution (or destination choice). This ordering is computationally convenient: trip distribution yields a set of origin-destination tables; the mode choice component then determines which modes of transport are used for each OD pair, with taking account of travel cost from the origin to destination as one of inputs. Arguably, this order also fits in a modelers' experience of real life in so far as on many occasions travelers' decide on their destinations first and then choose between transport modes.

The reverse order, implementing mode choice before destination choice, is less computationally straightforward: the destinations from production zones are unknown meaning that the generalized cost input for mode choice must be aggregated or composed at trip production level, instead of at trip (origin-destination) level. Due to the aggregation of cost inputs, the resultant base year validation outcomes tend to be less desirable than the destination choice first approach. As well, more importantly, the sequence affects elasticities of demand to cost/disutility changes and therefore, predicted future year demand.

The literature (UK Department of Transport 2014) supports the determination of the sequence of choice mechanisms according to the relative sensitivities (the lambdas of a logit model) of the choices to the generalised costs or disutilities of travel. If for example, mode choice has less sensitivity to cost or disutility changes than destination choice, then mode choice should be implemented before destination choice.

This case study compares the base year validation and future year sensitivity between the two different sequences, using to alternative approaches to estimating **Home Based Shopping** trips. The destination and mode choice models use a utility maximizing approach which is now commonly adopted in strategic level demand model, as in well described in the literature. The destination and mode choice forms are specified below.

3.2 Case study

3.2.1 Destination choice model form and parameters

The destination choice model is described by the formulation below:

$$P_{p,a} = \frac{exp(U_{p,a})}{\sum_{a' \in C}[exp\left(U_{p,a'}\right)]} \qquad \text{Equation (4)}$$

Where:
$P_{p,a}$ = Probability for trips from production zone p to choose destination zone a
a' = choice of an alternative zone
C = choice set that includes all feasible combinations of destination zones
$U_{p,a}$ = utility of choosing an alternative zone a for a given production zone p

The utility function is shown in the formulation below:

$$U_{pa} = \ln(A_a) + \lambda * \ GC_{pa} \qquad \text{Equation (5)}$$

Where: A_a = number of trip attraction at a destination zone a
= 6.05 * number of retail jobs + 1.47 * number of service jobs
GC_{pa} = Generalised cost from production zone p to destination zone a
λ = -0.173, sensitivity coefficient

3.2.2 Mode choice model form and parameters

The mode choice model is described by the formulation below:

$$P^m_{p,a} = \frac{exp(U^m_{p,a})}{\sum_{m' \in M}[exp\left(U^{m'}_{p,a}\right)]} \qquad \text{Equation (6)}$$

Where:
$P^m_{p,a}$ = Probability for a trip (from production zone p to choose destination zone a) to choose mode m
m' = choice of an alternative mode
M = choice set that includes all feasible combinations of modes
$U^m_{p,a}$ = Probability utility of choosing the mode m, for the trip from p to a

The utility value is calculated from the formulation below:

$$U^m_{p,a} = \lambda^m * \ GC^m_{p,a} + \alpha^m(Car\ Choice) + \beta^m(Car\ Captive) + \ ASC^m \qquad \text{Equation (7)}$$

Where: λ^m= sensitivity coefficient by mode- see Table 5
$GC^m_{p,a}$= Generalised cost from production zone p to destination zone a, for mode m
$Car\ Choice$ = dummy variable, 1 for Yes and 0 for No
$Car\ Captive$ = dummy variable, 1 for Yes and 0 for No
ASC^m = constant by mode

Table 5 Mode choice parameters

Parameter	Car driver	Car passenger	Walk /cycle	Public transport
λ	-0.124	-0.112	-0.873 ^	-0.029
α	2.296	1.718	1.570	-
β	-	-	1.162	-
ASC	-	-0.735	-0.193	-2.115

^ walk distance in kilometer is used as generalised cost input for the walk/cycle mode

Alternative 1 – Mode choice after destination choice

In application, the estimation of shopping trips by mode is done using the following steps:

- For each origin-destination trip, calculate the utility value for each mode according to Equation (7). Note that all the input variables are known: the generalised costs by mode are skimmed from network assignments for each OD movement and the Car Availability levels are known from the household and person information input to the model.
- Calculate the probability of each mode by apply the logit choice formulation of Equation (6).
- Aggregate the results over all origin-destination pairs, , and compare the observed (reported) and modelled mode shares to test the validation performance of the model

Table 6 shows the overall, aggregated modelled mode shares compared with observed values at a broad sector to sector level, whilst Figure 5 shows the performance of the model compared with observed trips. At both individual record level and broad sector SA2 to SA2 level the validation performance of the mode share is reasonable, noting that public transport has a small observed mode share (less than 2%), and the observed versus modelled discrepancy is largest for this mode.

Table 6 Overall mode shares, approach 1

Mode	Observed	Modelled
Car driver	65.6%	67.5%
Car passenger	25.6%	25.5%
Walk/Cycle	6.8%	5.6%
Public transport	2.0%	1.4%
Total	100%	100%

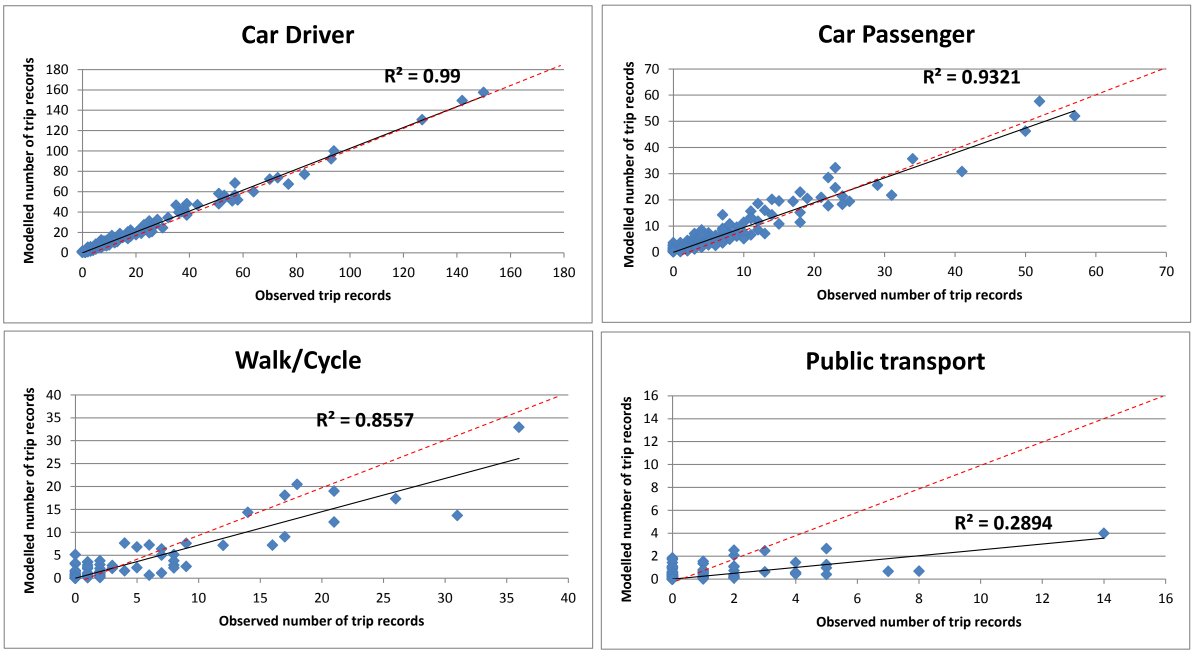

Figure 5 Observed and modelled trips, compared at SA2 to SA2 level, Alternative 1

Although the model validation performance is reasonable, Alternative 1 produces counter-intuitive changes in demand in response to cost changes, as shown in the example presented in Table 3**.** In the example, 1000 trips from the production zone A travel to two destination zone D1 and D2. The two destination zones have identical land use features and same travel costs from the production zone, as shown in Table 7. Therefore, the modelled destination and mode shares to each alternative destination zone are identical, as shown in Table 4.In the future year scenario, the car generalised cost to the destination D1 is increased from 30 to 40 minutes, reflecting deteriorating congestion from zone A to D1, whilst all other inputs remain unchanged. In the singly constrained application, set out in Table 8 (step 2.3), the outcome forecast of destination and mode causes a cause a counter-intuitive decrease in public transport trips to D1 (from 22 to 18 trips) in a single constrained model, as shown Table 8.

The deterioration of road congestion along a corridor would not be expected to reduce the public transport patronage as in this example, public transport travel costs remain unchanged. In application this issue can be 'solved' by also constraining the trips to zonal attraction totals, which is undertaken in a double constrained model, as shown in Step 2.5. However, this 'solution' disguises the counter-intuitive effect and dampens the elasticity of demand, which is illustrated by the comparison with Alternative 2

Table 7 Example model inputs. Alternative 1.

			Model input, base year		Model input, future year	
Variable for utility formulation		coefficient	Destination zone - D1	Destination zone - D2	Destination zone - D1	Destination zone - D2
For destination zone D	Number of trip attraction	-	500	500	500	500
	Ln(attraction)	1.00	6.2	6.2	6.2	6.2
	Generalised cost - weighted average all modes	-0.173	33	33	***44***	33
For car driver	Car generalised cost (min)	-0.124	30	30	***40***	30
	Car choice or not (1 for Y, 0 for N)	2.296	1	1	1	1
For car passenger	Car generalised cost (min)	-0.112	30	30	***40***	30
	Car choice or not (1 for Y, 0 for N)	1.718	1	1	1	1
	Constant	-0.735	1	1	1	1
For walk/cycle	Walk/cycle time (min)	-0.873	20	20	20	20
	Car choice or not (1 for Y, 0 for N)	1.570	1	1	1	1
	Car captive or not (1 for Y, 0 for N)	1.162	0	0	0	0
	Constant	-0.193	1	1	1	1
For public transport	PT generalised cost (min)	-0.029	70	70	70	70
	Constant	-2.115	1	1	1	1

Table 8 Example of Alternative 1 implementation

		Base year		Future year	
1. Implement destination choice		**D1**	**D2**	**D1**	**D2**
Calculate Utility value for each destination zone		0.49	0.49	-1.41	0.49
Take the expontentialed value of the utility		1.64	1.64	0.24	1.64
Calculate the probability of each destination zone		50%	50%	13%	87%
Calculate the number of trips to each destination zone		500	500	129	871
2. Implement mode choice		**D1**	**D2**	**D1**	**D2**
2.1 - Calculate utility value by mode	Car driver	-1.44	-1.44	-2.68	-1.44
	Car pax	-2.36	-2.36	-3.48	-2.36
	Walk/Cycle	-16.08	-16.08	-16.08	-16.08
	PT	-4.16	-4.16	-4.16	-4.16
2.2 - Take the expontentialed value of the utility	Car driver	0.237	0.237	0.068	0.237
	Car pax	0.094	0.094	0.031	0.094
	Walk/Cycle	0.000	0.000	0.000	0.000
	PT	0.016	0.016	0.016	0.016
	Sum	0.347	0.347	0.115	0.347
2.3 - Calculate the probability	Car driver	68%	68%	60%	68%
	Car pax	27%	27%	27%	27%
	Walk/Cycle	0%	0%	0%	0%
	PT	4%	4%	14%	4%
	Sum	100%	100%	100%	100%
2.4 - Calculate the number of Trips	Car driver	342	342	77	595
	Car pax	136	136	35	236
	Walk/Cycle	0	0	0	0
	PT	22	22	***18***	***39***
	Sum	500	500	129	871
2.5 - Constrain to trip attractions - if the double constrained model is adopted	Car driver	342	342	298	342
	Car pax	136	136	134	136
	Walk/Cycle	0	0	0	0
	PT	22	22	***68***	22
	Sum	**500**	**500**	**500**	**500**

Alternative 2 – Mode choice before destination choice

The model application processes are similar to the Alternative 1. The only difference occurs on step 1 where the utility values are composed at the trip production level because the destination zones remain unknown before the destination choice. For each observed trip record, calculate the utility value for each mode according to Equation (8). Given the destination zones are unknown before the destination choice, for each record the utility from its production zones to all possible destination zones (zones a1 to an) are calculated, and their logsum value is used as the composite utility value the trip production level, as shown in the formulation below.

$$U_p^m = \ln[\exp(U_{p,a1}^m) + \exp(U_{p,a2}^m) + \cdots + \exp(U_{p.an}^m)] \qquad \text{Equation (8)}$$

Where: U_p^m= utility value for the trips from production zone p, using mode m

$U_{p,a}^m$= utility value for the trips from production zone p to attraction zone a, using mode m

This is followed by calculating the probability of each mode by apply the logit choice formulation of Equation (6). Table 9 reports the overall mode share comparisons between modelled and observed whilst Figure 6 shows the comparison at SA2 level. It shows that the overall mode shares are still close to the observed values. At the SA2 to SA2 level, Approach 2 produces slightly worse fitness than Approach 1 (slightly lower R square), in particular for the minor modes such as walk/cycle and public transport.

However, Approach 2 produces more sensible and sensitive outcomes in response to the model input changes discussed in the section above. As shown in Table 10, for the outcomes of the single constrained model (step 2.3), the PT demand increases to destination D1 (instead of decreasing in Approach 1), reflecting the shift of car users to public transport. If the model is double constrained, as shown in 2.4, the resultant PT trips to D1 is 106, significantly larger than the 68 trips in Approach 1, meaning that the Approach 2 produces more elastic responses in demand as cost changes.

Table 9 Overall mode shares, Approach 2

Mode	Observed	Modelled
Car driver	65.6%	67.3%
Car passenger	25.6%	24.0%
Walk/cycle	6.8%	7.5%
Public transport	2.0%	1.2%
Total	100%	100%

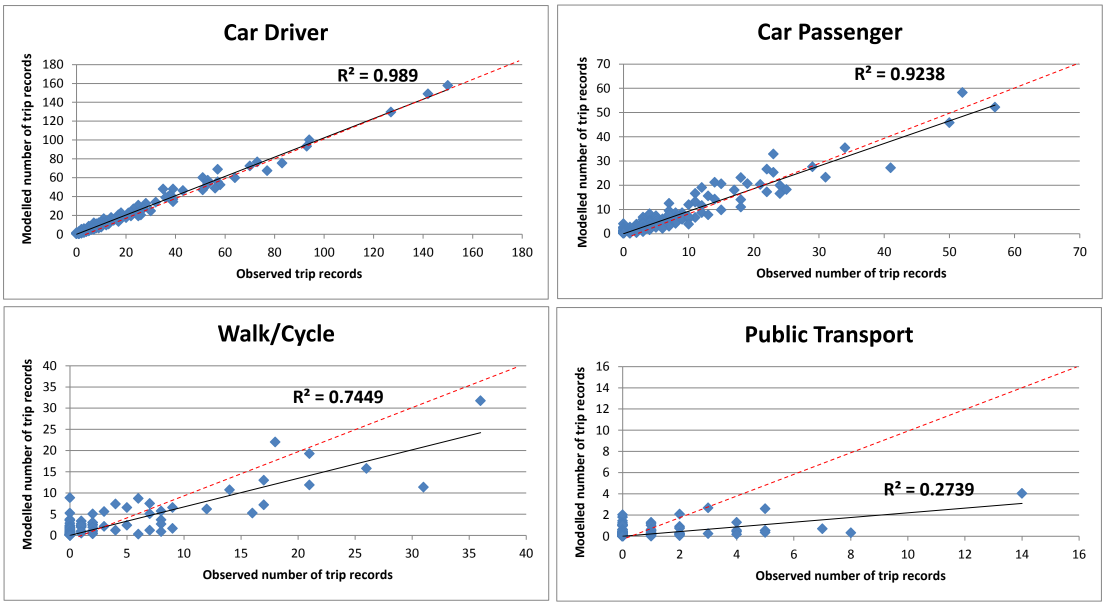

Figure 6 Observed and modelled trip records compared, at SA2 to SA2 level, Approach 2

Table 10 Example of Alternative 2 implementation

		Base year		Future year	
1. Implement Mode choice		**D1**	**D2**	**D1**	**D2**
1.1 - Calculate utility value by mode	Car driver	-1.44	-1.44	-2.68	-1.44
	Car pax	-2.36	-2.36	-3.48	-2.36
	Walk/Cycle	-16.08	-16.08	-16.08	-16.08
	PT	-4.16	-4.16	-4.16	-4.16
1.2 - Take the expontentialed value of the utility	Car driver	0.237	0.237	0.068	0.237
	Car pax	0.094	0.094	0.031	0.094
	Walk/Cycle	0.000	0.000	0.000	0.000
	PT	0.016	0.016	0.016	0.016
1.3 - Calculate the logsum of the utility values for D1 and D2	Car driver	-0.75		-1.19	
	Car pax	-1.67		-2.08	
	Walk/Cycle	-15.39		-15.39	
	PT	-3.47		-3.47	
1.4 - Take the expontentialed value of the composite utilities	Car driver	0.47		0.31	
	Car pax	0.19		0.12	
	Walk/Cycle	0.00		0.00	
	PT	0.03		0.03	
	Sum	0.69		0.46	
1.5 - Calculate the probability of each mode	Car driver	68%		66%	

		Base year		Future year	
1. Implement Mode choice		**D1**	**D2**	**D1**	**D2**
	Car pax	27%		27%	
	Walk/Cycle	0%		0%	
	PT	4%		7%	
	Sum	100%		100%	
1.6 - Calculate the number of Trips by mode	Car driver	684		662	
	Car pax	271		271	
	Walk/Cycle	0		0	
	PT	45		68	
	Sum	1000		1000	
2. Implement Destination choice		**D1**	**D2**	**D1**	**D2**
2.1 - Calculate utility value by destination - note that the cost input is now distinguished for each mode (instead of using blended average)	Car driver	0.49	0.49	-1.41	0.49
	Car pax	1.01	1.01	-0.72	1.01
	Walk/Cycle	2.75	2.75	2.75	2.75
	PT	-5.92	-5.92	-5.92	-5.92
2.2 - Calculate the probability for each destination (by mode)	Car driver	50%	50%	13%	87%
	Car pax	50%	50%	15%	85%
	Walk/Cycle	50%	50%	50%	50%
	PT	50%	50%	50%	50%
2.3 - Calculate the number of trips to each destination (by mode)	Car driver	342	342	86	576
	Car pax	136	136	41	230
	Walk/Cycle	0	0	0	0
	PT	22	22	***34***	***34***
	Sum	500	500	160	840
2.4 - Constrain to trip attractions - if the double constrained model is adopted	Car driver	342	342	267	343
	Car pax	136	136	127	137
	Walk/Cycle	0	0	0	0
	PT	22	22	***106***	20
	Sum	500	500	500	500

4. Conclusions

This paper has presented modelling case studies comparing validation performance and the performance in forecasting. We have shown that achieving acceptable validation performance is not in itself an indicator of a reliable model.

Overfitting measures can reduce a model's responsiveness to the changes in model inputs, undermining the reliability of forecasting outcomes. The alternative models which removing the over-fitting measures may have less desirable validation outcomes, but can produce more responsive and sensible outcomes in forecasting. Our conclusion is that recognizing relative underperformance in the base year validation in exchange for more reasonable forecasting outcomes may be preferred to overfitting to observed data. In practice, this may means that model development effort improving the base year might be better spent testing the model's performance in forecasting, critically examining the performance of the model in forecasting together with developing a range of forecasts and sensitivity tests, which explore the uncertainties attached to each of the input assumptions.

References

Bain, R. (2015). PPP Lessons from Europe. *AITPM National Conference*. Brisbane: AITPM.

Department of Transport (2014) Forecasting and Uncertainty, TAG Unit M4
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/427130/TAG_Unit_M4_Forecasting_and_Uncertainty_November2014.pdf

Hall, Randolph (2003). *Handbook of Transportation Science*

Ortuzar and Willumsen (2011) Modelling Transport, 4th Edition

Presenters Bio

Tracey Pershouse is an Associate Director with AECOM and has more than twenty years experience in demand forecasting and transport modelling, working in Australia, Hong Kong and the United Kingdom. She has developed and applied transport models to appraise investments in highway, public transport as well as active modes. She currently works in AECOM's Brisbane office, developing and applying models in for a number of Queensland's regional cities, as well as urban models in Sydney and Brisbane.